Environmental Research Laboratories

Air Resources
Atmospheric Turbulence and Diffusion Laboratory

Oak Ridge, Tennessee

September 1974

A METHOD FOR EVALUATING THE ACCURACY
OF AIR POLLUTION PREDICTION MODELS

Carmen J. Nappo, Jr.

ATDL Contribution File No. 100

U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

A Method for Evaluating the Accuracy of Air Pollution Prediction Models

Carmen J. Nappo, Jr.

Air Resources
Atmospheric Turbulence and Diffusion Laboratory
National Oceanic and Atmospheric Administration
Oak Ridge, Tennessee

(Published in the proceedings of the Symposium on Atmospheric Diffusion and Air Pollution held September 9-13, 1974 in Santa Barbara, California.)

A METHOD FOR EVALUATING THE ACCURACY OF AIR POLLUTION PREDICTION MODELS

Carmen J. Nappo, Jr.

Air Resources
Atmospheric Turbulence and Diffusion Laboratory
National Oceanic and Atmospheric Administration
Oak Ridge, Tennessee

1. INTRODUCTION

The accuracy of an air pollution prediction model can be evaluated only by measuring its ability to reproduce an air pollution episode. This ability is often measured by comparing the time changes of the predicted with the observed air pollution concentrations at several monitoring stations within the prediction area. Often, the predicted and observed time-average concentrations are compared and/or the temporal correlation coefficients are formed between these quantities at each station. This rather one-dimensional view has resulted in a controversy over which of the current urban air pollution prediction models is best. Table 1 shows a typical result of such an evaluation. The column labeled "Model" lists the references as well as the numerical technique used, i.e. trajectory, particle-in-cell, etc. All of these models predict carbon monoxide, CO, concentration from extended area sources and all but the model by MacCracken, et al. (1971) are applied to the Los Angeles basin. The model by MacCracken, et al. is applied to the San Francisco Bay area. In the column labeled "Average Temporal Correlation," the station average of the temporal correlation coefficient formed for each model is listed. In cases where several predictions were performed, the average of the correlations from all predictions is used. The column labeled "Computation Time," estimates the computer time, in minutes, required by each model for a 24-hour forecast using an IBM 360/65 machine. Finally under "Computer Cost," the approximate cost for this 24-hour prediction, in dollars, is presented.

One obvious result of such an evaluation is that although many of the models are roughly of equal accuracy, there is a great disparity in their complexity and operating cost. This is the basis for the above mentioned controversy. In this discussion, a more comprehensive method of evaluation is proposed and applied to the CO air pollution models listed in Table 1. The method in essence takes into account the spatial variability as well as the temporal variability of the prediction. The following are a few of the results observed: 1) models which were previously regarded accurate on the basis of their time correlations, are not so accurate when the comprehensive evaluation is used; 2) the detailed modeling of vertical diffusion is of little significance in determining ground concentrations of CO; and 3) model predictions are sensitive mostly to the source emissions.

TABLE 1

MODEL EVALUATION BASED ON TEMPORAL CHARACTERISTICS

Model	Average Temporal Correlation Coefficient	Computer Time for 24 Hour Prediction (min)	Computer Cost for 24 Hour Prediction (dollars)
MacCracken *et al* (1971) multi-box	0.37	106	350
24 Hour Persistence	0.47	None	None
Roth *et al* (1971) primitive equation	0.52	60	200
Hanna (1973) ATDL simple model	0.60	None	None
Sklarew *et al* (1972) particle-in-cell	0.65	49	160
Pandolfo and Jacobs (1973) primitive equation	0.66	20	70
Reynolds *et al* (1973) primitive equation	0.73	30	100
Eschenroeder *et al* (1972) trajectory	0.73	15	50
Lamb and Neiburger (1971) trajectory	0.90	35	115

2. THE METHOD

If a prediction model is to accurately reproduce an air pollution episode, it must reproduce at each monitoring station the observed time-varying pollution concentration, and reproduce at each monitoring time the observed space-varying pollution pattern. The degree with which the observed time-varying concentration is reproduced is measured in part by the station (spatial) average of the temporal correlation coefficient formed at each station, $\overline{R(t)}^s$. This term can be regarded as the measure of the model's ability to reproduce the observed temporal trends of air pollution over the whole network of monitoring stations. In the same manner, the degree with which the observed spatial trends of air pollution are reproduced over the entire prediction period is measured by the time average of the correlation coefficient formed between the predicted and observed patterns of the concentration isopleths at each monitoring time, $\overline{R(s)}^t$. The temporal and spatial correlation coefficients are discussed in the Appendix. Ideally, $\overline{R(t)}^s$ and $\overline{R(s)}^t$ would equal unity. Often, however, there is insufficient spatial resolution of the data to form the isopleths needed to determine $\overline{R(s)}^t$. Instead, at each observation time, a correlation coefficient is formed using the predicted and observed pollution concentrations at all monitoring stations. It is expected that this term in quality, at least, reflects the correlation of the observed and predicted concentration patterns.

As indicated above, $\overline{R(s)}^t$ and $\overline{R(t)}^s$ do not completely measure a model's accuracy. The ability of a model to reproduce the time and space varying amounts of air pollution must also be measured. This can be done in the following way. At each monitoring station, the time average of the ratio of predicted to observed concentration is formed; these time-averaged ratios are then averaged over all stations and called $\overline{r(t)}^s$. Next, at each monitoring time, the space average of the ratio of predicted to observed concentration is formed; these are averaged over all monitoring times and called $\overline{r(s)}^t$. The formation of these averaged ratios is explained in the Appendix. Because of the way these averages are formed, $\overline{r(t)}^s$ equals $\overline{r(s)}^t$ (see Appendix), and this number represents the totally averaged ratio of the predicted to observed air pollution concentration. If a model reproduces exactly the air pollution averaged over space and time (e.g. the 24-hour averaged air pollution in the Los Angeles basin), then for the model $\overline{r(s)}^t = \overline{r(t)}^s = 1$. However, while these averaged ratios are equal, their standard deviations are not, and these deviations represent the error or variance contained in the models prediction of the time-and space-averaged amounts of the air pollution. Let the space average of the variance in the time-averaged prediction be designated by $\overline{\sigma(t)}^s$, and the time average of the variance in the space-averaged prediction be designated by $\overline{\sigma(s)}^t$.

The evaluation of the accuracy of an air pollution prediction model requires the formation of the following quantities:

1 $\overline{R(t)}^s$, $\overline{R(s)}^t$, $\overline{r(t)}^s \pm \overline{\sigma(t)}^s$, and $\overline{r(s)}^t \pm \overline{\sigma(s)}^t$.

The ideal model would have

2 $$\overline{R(t)}^s = \overline{R(s)}^t = 1$$

and

3 $$\overline{r(t)}^s \pm \overline{\sigma(t)}^s = \overline{r(s)}^t \pm \overline{\sigma(s)}^t = 1.$$

3. ILLUSTRATION OF THE METHOD

As an illustration of this evaluation method, the above quantities are formed for each of the CO prediction models listed in Table 1. The plot of the space average of the temporal correlation coefficient, $\overline{R(t)}^s$, versus the time average of the spatial correlation coefficient, $\overline{R(s)}^t$, is shown in Figure 1 for each model. It is seen that the models reproduce the temporal trends better than the spatial

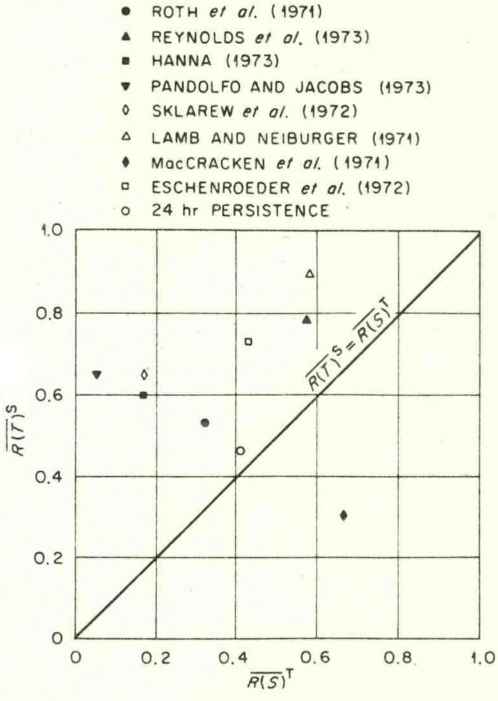

Figure 1. $\overline{R(t)}^s$ versus $\overline{R(s)}^t$
Average result for each model tested.

trends, with the exception of the San Francisco Bay area model (MacCracken, et al.). The differences in the ability of the models to reproduce the spatial trends is also quite obvious. In Figure 2, the space average of the time-averaged ratios of the predicted to observed CO concentration, $\overline{r(t)}^s$, is plotted against the time average of the space-averaged ratios of the predicted to observed concentrations, $\overline{r(s)}^t$, with accompanying error bars representing the standard deviations $\overline{\sigma(t)}^s$ and $\overline{\sigma(s)}^t$. Comparing Figures (1) and (2), several significant features are observed. First of all, while Hanna's application of the ATDL simple air pollution model reproduces the observed trends of pollution as well as the complicated particle-in-cell model (Sklarew, et al.), there is a great difference in their ability to predict the amounts of air pollution, with the ATDL simple model seriously over-estimating the observed concentrations. Secondly, the trajectory model of Lamb and Neiburger which from Table 1 and Figure (1) might be judged the best of the models, seriously underpredicts the observed CO concentration. Finally, while the accuracy of the temporal and spatial trend prediction of the Systems Application 1973 model (Reynolds, et al.) is greatly increased over its original 1971 form (Roth, et al. (1971)), there is no correspondingly large change in ability to predict the amount of CO.

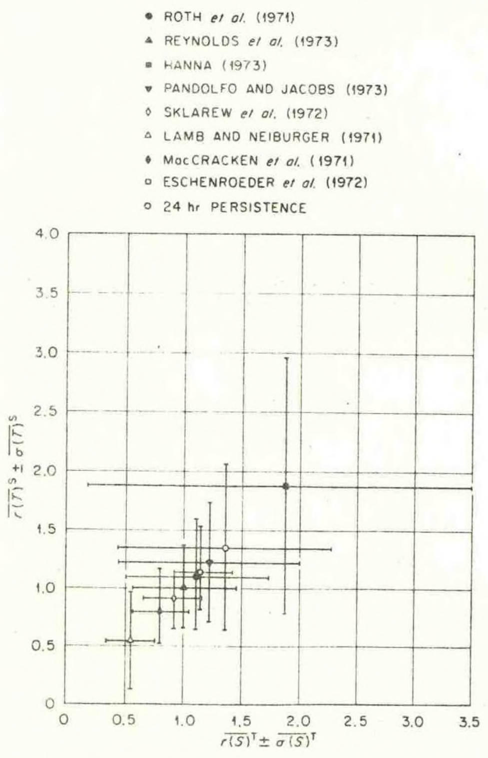

Figure 2. $\overline{r(t)}^s \pm \overline{\sigma(t)}^s$ versus $\overline{r(s)}^t \pm \overline{\sigma(s)}^t$
Average result for each model tested.

In Figure (3), $\overline{\sigma(t)}^s$ is plotted against $\overline{\sigma(s)}^t$. Again we see that while Hanna (1973) and Sklarew, et al. (1972) reproduce the trends equally as well, there is a great difference in the variances of the predicted amounts of air pollution. We also see that the primitive equation models have greater error (variance) in their spatial predictions than in their temporal predictions, i.e. $\overline{\sigma(s)}^t > \overline{\sigma(t)}^s$, while the reverse is true for the trajectory, particle-in-cell, and box models.

4. RESULTS OF THE EVALUATION

From the above illustration of the evaluation of the evaluation method, two results follow. First of all, it seems clear that the detail of the vertical diffusion calculations is not very significant in determining a model's accuracy. Note that Lamb and Neiburger's model has no vertical detail in the distribution of pollution concentration. The Systems Application 1971 model had ten levels in the vertical while the 1973 version has only five. The model of Eschenroeder, et al. (1972) also has five levels, while the ATDL simple model assumes a gaussian distribution of pollutants in the vertical. This observation is further confirmed by Reynolds, et al. (1973) who compared a two-dimensional (no vertical diffusion) model with their five-layer model, and found that "differences in prediction were generally (though not always)

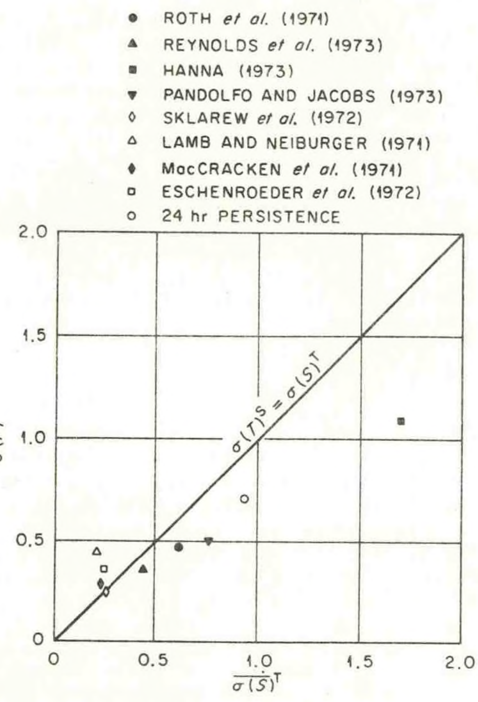

Figure 3. $\overline{\sigma(t)}^s$ versus $\overline{\sigma(s)}^t$
Average result for each model tested.

rather small." They go on to recommend that "the question of dimensionality be further explored in future studies of model sensitivity."

The second result is that model sensitivity and accuracy are dependent mostly on the degree of detail of the source emissions inventory. This is quite evident in comparing the Systems Application 1971 and 1973 models and their respective results. The 1973 meteorological model has been simplified ("detuned") with respect to its 1971 form. For example, the original vertical resolution is halved, and a less accurate finite differencing scheme is used for the horizontal advection together with a less detailed wind field. However, in the 1973 model a much more detailed source emissions inventory is utilized. This appears to have more than compensated for the "detuning" of the earlier meteorological model because the resulting overall accuracy has been improved.

5. CONCLUSION

A method for evaluating the accuracy of air pollution models has been proposed. The method measures a model's ability to reproduce the observed spatial and temporal trends of air pollution, as well as the observed spatial and temporal amounts of air pollution and their respective errors. This method was illustrated by evaluating several models of CO air pollution. Initial results of this evaluation are that: 1) models previously regarded as quite accurate are infact, less accurate when comprehensively evaluated; 2) the Los Angeles Basin models reproduce the observed temporal trends of air pollution better than observed spatial trends, while the reverse is true for a San Francisco Bay area model; 3) primitive equation models have greater variance in their spatial predictions of air pollution concentrations than in their temporal predictions, while the reverse is true for trajectory, particle-in-cell and box models; 4) the detailed calculation of vertical diffusion does not appear significant in the prediction of air pollution ground concentrations; and 5) a model's accuracy is sensitive mostly to the degree of detail of the source emission inventory.

APPENDIX: FORMATION OF THE STATISTICS

Let the predicted and observed concentrations at the t'th monitoring time and the s'th monitoring station be given by $P_{s,t}$ and $O_{s,t}$ respectively. The temporal correlation coefficient is formed at each monitoring station by

$$\mathcal{R}_s(t) = \frac{\sum_{t=1}^{T} \mathcal{P}_{s,t}\, \mathcal{O}_{s,t}}{\sqrt{\sum_{t=1}^{T} (\mathcal{P}_{s,t})^2}\ \sqrt{\sum_{t=1}^{T} (\mathcal{O}_{s,t})^2}} \quad (1)$$

where T is the number of monitoring times and

$$\mathcal{P}_{s,t} = P_{s,t} - \overline{P_{s,t}}^{\,t} \quad (2)$$

$$\mathcal{O}_{s,t} = O_{s,t} - \overline{O_{s,t}}^{\,t} \quad (3)$$

The overbar denotes a time average. The spatial average of the temporal correlation coefficient, $\overline{R(t)}^{s}$, is given by

$$\overline{\mathcal{R}(t)}^{\,s} = \frac{1}{N} \sum_{s=1}^{N} \mathcal{R}_s(t) \quad (4)$$

where N is the number of monitoring stations.

The spatial correlation coefficient is formed at each monitoring time by

$$\mathcal{R}_t(s) = \frac{\sum_{s=1}^{N} \mathcal{P}'_{s,t}\, \mathcal{O}'_{s,t}}{\sqrt{\sum_{s=1}^{N} (\mathcal{P}'_{s,t})^2}\ \sqrt{\sum_{s=1}^{N} (\mathcal{O}'_{s,t})^2}} \quad (5)$$

where

$$\mathcal{P}'_{s,t} = P_{s,t} - \overline{P_{s,t}}^{\,s} \quad (6)$$

$$\mathcal{O}'_{s,t} = O_{s,t} - \overline{O_{s,t}}^{\,s} \quad (7)$$

The overbar denotes a space or station average. The temporal average of the spatial correlation coefficient is given by

$$\overline{\mathcal{R}(s)}^{\,t} = \frac{1}{T} \sum_{t=1}^{T} \mathcal{R}_t(s) \quad (8)$$

The time average and the space average of the ratio of the predicted to observed concentrations is given respectively by

$$r_s(t) = \frac{1}{T} \sum_{t=1}^{T} (P_{s,t}/O_{s,t}) \quad (9)$$

and

$$r_t(s) = \frac{1}{N} \sum_{s=1}^{N} (P_{s,t}/O_{s,t}) \quad (10)$$

The space average of the time-averaged ratios is

$$\overline{r(t)}^{s} = \frac{1}{N}\sum_{s=1}^{N} r_s(t) \tag{11}$$

The time average of the space-averaged ratios is given by

$$\overline{r(s)}^{t} = \frac{1}{T}\sum_{t=1}^{T} r_t(s) \tag{12}$$

We can easily show that $\overline{r(t)}^{s} = \overline{r(s)}^{t}$ by using (9) to (12) as follows.

$$\begin{aligned}
\overline{r(t)}^{s} &= \frac{1}{N}\sum_{s=1}^{N} r_s(t) \\
&= \frac{1}{N}\frac{1}{T}\sum_{s=1}^{N}\sum_{t=1}^{T} (P_{s,t}/O_{s,t}) \\
&= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{s=1}^{N} (P_{s,t}/O_{s,t}) \\
&= \frac{1}{T}\sum_{t=1}^{T} r_t(s) \\
&= \overline{r(s)}^{T}
\end{aligned}$$

ACKNOWLEDGEMENT

This work was performed under an agreement between the National Oceanic and Atmospheric Administration and the U. S. Atomic Energy Commission.

REFERENCES

Eschenroeder, A. Q., J. R. Martinez and R. A. Nordsieck, 1972: Evaluation of a diffusion model for photochemical smog simulation. Final Rpt. No. 68-02-0336 for the Environmental Protection Agency. General Res. Corp., Santa Barbara, Cal.

Hanna, S. R., 1973: Urban air pollution models - why? ATDL contribution file no. 83, Atmospheric Turbulence and Diffusion Lab., Oak Ridge, Tennessee.

Lamb, R. G. and M. Neiburger, 1971: An interim version of a generalized urban air pollution model. Atmos. Environ., No. 4, Vol. 5, pp 239-264.

MacCracken, M. C., T. V. Crawford, K. R. Peterson, and J. B. Knox, 1971: Development of a multi-box air pollution model and initial verification for the San Francisco Bay area. Lawrence Radiation Lab. No. UCRL-73348, Livermore, Cal.

Pandolfo, J. P. and C. A. Jacobs, 1973: Tests of an urban meteorological-pollutant model using CO validation data in the Los Angeles metropolitan area. Vol. 1, EPA Rpt. No. R4-730-025A, The Center for the Environ. and Man, Inc., Hartford, Conn.

Reynolds, S. D., M. K. Liu, T. A. Hecht, P. M. Roth, and J. H. Seinfeld, 1973: Further development and evaluation of a simulation model for estimating ground level concentrations of photochemical pollutants. Final Rpt. EPA Contract 68-02-0339, Systems Applications, Inc., Beverly Hills, Cal.

Roth, P. M., S. D. Reynolds, P. J. W. Roberts, and J. H. Seinfeld, 1971: Development of a simulation model for estimating ground level concentrations of photochemical pollutants. Final Rpt. 71-SA1-21, Systems Applications, Inc., Beverly Hills, Cal.

Sklarew, R. C., A. J. Fabrick, and J. E. Prager, 1971: A particle-in-cell method for numerical solution of the atmospheric diffusion equation, and applications to air pollution problems. Final Rpt. 3SR-844, Systems, Science and Software, La Jolla, Cal.